Discrete Applied Mathematics 157 (2009) 571–582

Contents lists available at ScienceDirect

Discrete Applied Mathematics

journal homepage: www.elsevier.com/locate/dam

Feature controlled adaptive difference operators

Peter Veelaert *, Kristof Teelen

Faculty of Applied Engineering Sciences, University College Ghent, Schoonmeersstraat 52, B9000 Ghent, Belgium

ARTICLE INFO

Article history:
Received 14 May 2007
Received in revised form 22 November 2007
Accepted 15 May 2008
Available online 7 July 2008

Keywords:
Difference operators
Groebner basis
Local feature detectors

ABSTRACT

Differential operators are essential for many image processing applications which require the computation of typical characteristics of continuous surfaces, as e.g. tangents, curvature, flatness, shape descriptors. We propose to replace differential operators by the combined action of sets of feature detectors and locally adaptive difference operators, resulting in a more accurate computation of the required derivatives in each pixel neighborhood. Both the set of feature detectors and the set of difference operators have a rigid mathematical structure, which is described by a set of Groebner bases for each class of fitting functions. This representation allows a systematic description of the hierarchical structure with ordering relations for all different function classes. The explicit computation of fitting functions is avoided by our technique and replaced by a function classification process. A set of simple local feature detectors is used to find the class of fitting functions which locally yields the best approximation for the digitized image surface. By a systematic optimization process, we determine for each fitting function class a difference operator which is an optimal approximation for a particular differential operator. As an example, we describe how to compute the best discrete approximation for the Laplacian differential operator in each pixel neighborhood and illustrate how the Laplacian of Gaussian edge detection method can benefit from these results.

© 2008 Elsevier B.V. All rights reserved.

1. Introduction

A delicate and often recurring problem in digital image processing is the application of operators from differential geometry to discrete representations of curves and surfaces. The discrete surface is often formed by the discrete intensities of a digital image. For continuous surfaces well-defined differential operators are used to compute standard functions such as curvature, tangent, or shape operators. These differentials cannot be applied directly to digitized surfaces or digitized curves. In image processing, this obstacle is often circumvented by fitting a continuous function $\tilde{g}$ to the discrete function f in some local neighborhood of point, and to apply the differential operator to the continuous function.

Haralick [1], Langridge [2], Fleck [3], Karabassis and Spetsakis [4] all discuss possible methods to find the best possible fitting functions and difference operators to compute derivatives for a given application. For specific classes of images, and with appropriate fitting functions, the results are often good. One major advantage of fitting functions is their relative insensitivity to image noise, that is, quantization noise as well as camera noise, which form a serious problem when computing derivatives. The use of fitting functions, however, also has considerable drawbacks. First, computing fitting functions is time consuming, especially for high-order fitting functions, which are necessary for computing high-order derivatives. Second, several important choices must be made in advance, such as the size and shape of the neighborhood in which the approximation takes place, and the form and order of the fitting function. These choices have a considerable, and sometimes undesirable, influence on the result. For example, in an image the use of quadratic fitting functions may

* Corresponding author. Tel.: +32 9 248 88 26; fax: +32 9 243 87 77.
E-mail addresses: peter.veelaert@hogent.be (P. Veelaert), kristof.teelen@hogent.be (K. Teelen).

0166-218X/$ – see front matter © 2008 Elsevier B.V. All rights reserved.
doi:10.1016/j.dam.2008.05.024

result in over fitting where the intensity varies almost linearly, and under fitting where there are sharp edges. A better solution would be to examine in each pixel several neighborhoods as well as several classes of fitting functions, and to select the most appropriate class and neighborhood before computing the differential. Such a systematic analysis requires huge computational resources, however, and is therefore seldom used in image processing.

Instead of computing a continuous approximation, we propose to digitize the differential operator locally, i.e. to replace the differential at each point by the most appropriate difference operator at that point. The difference operator is applied directly to the discrete function. We actually avoid the problem of computing a fitting function by simply verifying whether a digitized function has the right features or characteristics to be assigned to one of the function classes. For each class there is an optimal difference operator, which gives the smallest error when it replaces the differential operator. The advantages of using fitting functions are not lost, however. One of the main results in this paper is that the outcome of an optimal difference operator is very close to the outcome of a differential operator combined with a fitting function. The outcome of an adaptive difference operator is certainly superior to what can be obtained by using a single difference operator for an entire image, as a single operator will only be optimal for one particular class of functions.

Adaptive difference operators that skip the approximation step have additional benefits. The computation time is considerably reduced, so that more than one neighborhood and/or more than one class of fitting functions can be considered, before selecting the best operator. The fitting classes are not chosen ad hoc, but they can be chosen from a logical hierarchy of classes. The feature detection aspect arises in a natural way within this hierarchy. We will show that the selection of the optimal difference operator as well as the choice of appropriate feature detectors both fit within a rigid mathematical framework, which involves ideals of difference operators.

There are numerous applications of differential and difference operators in image processing. Lachaud et al. [5] discuss how to estimate the tangent of a digital curve. Lindeberg [6] discusses how to define discrete derivative approximations for the computations of multi-scale low-level feature extraction, and their use in edge detection. Gunn [7] and Demigny and Kamlé [8] consider discrete versions of edge detection algorithms. The Laplacian of Gaussian (LoG) on different scales is used in several applications to extract edges or feature points. To illustrate the proposed approach, we show in this paper how a LoG edge detector can benefit from locally applying a feature controlled adaptive Laplacian difference operator.

This paper extends previous work on difference operators and feature detectors for discrete functions [9,10]. In Section 2, we show how to digitize the differential operator for different classes of fitting functions. Next, a decision chain for the practical computation of difference operators is introduced in Section 3, where the computation of the feature detection templates and the difference operators for the Laplacian is given as an example. Section 4 presents the Laplacian of Gaussian edge detector as a practical application to illustrate the benefits of adaptive operators. Finally, we conclude this paper in Section 5.

2. Digitizing differential operators

When choosing a class of fitting functions for a digitized function, we actually introduce a class of desirable function characteristics, which we shall call features. Simple feature detectors will examine how a digitized function behaves locally, that is, whether the function is either linear, quadratic, sufficiently smooth, symmetric, and so on. We will show that feature detection can be realized by the parallel verification of simple inequalities that must be satisfied by the function values. Depending on the detected features and the differential operator, there is an optimal difference operator. In this section we show how to select appropriate classes of fitting functions, feature detectors and optimal difference operators. We start by introducing the notations and the conventions used in this paper.

We introduce a continuous real function $\tilde{g} : \mathbb{R}^m \to \mathbb{R}$ to approximate a digitized function $f : \mathbb{Z}^m \to \mathbb{Z}$. To approximate the value of a differential at a point x_0, it is sufficient to approximate f in a finite subset $D \subset \mathbb{Z}^m$ containing x_0. For every $x \in D$, we have $f(x) \simeq \tilde{g}(x)$. We write $|f - \tilde{g}| < \epsilon$ as a shorthand for $|f(x) - \tilde{g}(x)| < \epsilon$ for all $x \in D$.

The shift operator σ^j is defined by $\sigma^j f(x) = f(x + j)$, for $x, j \in \mathbb{Z}^m$. The functional composition of shift operators is expressed as the multiplication of polynomials, i.e. $\sigma^j \sigma^k f = \sigma^{j+k} f$. A difference operator P can be represented as a polynomial in σ with non-negative, bounded exponents, that is $P = \sum p_j \sigma^j$, and $P \in \mathbb{R}[\sigma]$, the ring of polynomials in σ. We write $P\tilde{g} = 0$, as a shorthand for $\sum p_j \sigma^j \tilde{g}(x) = 0$, with $x, j \in \mathbb{Z}^m$. Similarly, $|Pf - P\tilde{g}| < \epsilon$, means that $|Pf(x) - P\tilde{g}(x)| < \epsilon$ for all x for which $Pf(x)$ is well defined, that is $(x+j) \in D$ for every non-vanishing coefficient p_j of the difference operator P. Since polynomial ideals are usually defined for polynomials with exponents that are non-negative, without loss of generality, we will assume that D contains only points with non-negative coordinates. Furthermore, to facilitate some of the derivations we assume that D is rectangular, although the results can be extended to non-rectangular neighborhoods.

An important concept is the ideal I of difference operators that is generated by a set of operators P_i. The ideal I consists of all operators $P = \sum_i S_i P_i$, where S_i are arbitrary polynomials in σ, i.e. $S_i \in \mathbb{R}[\sigma]$. The polynomials P_i form a generating subset for the ideal I, which is denoted as $I = \langle P_1, P_2, \ldots \rangle$. In Section 2.4, we explain how to compute a sufficient number of generating operators for the ideal.

Difference operators can be represented by templates (or kernels), which are convenient when illustrating two-dimensional operators. The two-dimensional operator $P = \sum p_j \sigma^j = \sum_{j_x, j_y} p_{j_x j_y} \sigma_x^{j_x} \sigma_y^{j_y}$, $j \in \mathbb{Z}^2$ is represented by the template:

$$\begin{array}{|c|c|c|c|} \hline p_{00} & p_{10} & p_{20} & \dots \\ \hline p_{01} & p_{11} & p_{21} & \\ \hline p_{02} & \dots & & \\ \hline \end{array} \tag{1}$$

We use the convention that the box at the upper left corner corresponds to p_{00}. Boxes with vanishing coefficients are either not drawn, or drawn as empty boxes.

A standard approach to find the value of a differential operator when applied to a discrete function f is to approximate the discrete function first with a continuous function $\tilde{g}$, and apply the differential operator to the continuous approximation. Although this procedure gives good results, it is not so often applied in image processing, due to the computational burden necessary to find good continuous approximations. The main idea presented in this paper is that one can obtain the same result without computing the continuous approximation itself. The proposed approximation of the differential consists of two steps. First, it is verified whether f can be approximated well by a certain class of fitting functions without actually calculating the best fitting function. Second, once the class of fitting functions is known, a difference operator known to be optimal for this class is applied directly to the discrete function f.

To be precise, let L be the differential operator that must be replaced by a difference operator. The goal is to select a class G of fitting functions $\tilde{g}$, and a difference operator $Q = \sum q_j \sigma^j$ that satisfies $Q\tilde{g} = L\tilde{g}$ for every $\tilde{g} \in G$. If $\tilde{g}$ is an approximation for f such that $\lfloor f - \tilde{g} \rfloor < \epsilon$, then, because all operators are linear, we have

$$|Qf - L\tilde{g}| < \epsilon \sum |q_j|. \tag{2}$$

Hence, the difference operator Q is a good approximation for the differential operator L, provided G contains at least one good approximation $\tilde{g}$ for f. It is important, however, to understand that the total error $\epsilon \sum |q_j|$ is due to two successive approximations. The error ϵ arises when f is replaced by some continuous fitting function $\tilde{g}$, a step which cannot be avoided if we want to compute differentials. Second, the additional error $\sum |q_j|$ is due to the replacement of the differential operator L by a difference operator Q to speed up the computation, that is, by avoiding the explicit computation of the continuous approximation. One may argue that when there is a sufficient amount of computational resources, this additional error is unnecessary. Even then, however, a speedup in computation time always has the advantage that there will be more time to look for the best class of fitting functions, that is, the class which yields a small value for ϵ. This advantage may outweigh the additional error $\sum |q_j|$.

Clearly, the choice of classes of fitting functions has important consequences. The criteria that will be used to select appropriate classes are the following:

(1) The class G of fitting functions must be *shift-invariant*, that is if $\tilde{g}(x, y) \in G$, then $\tilde{g}(x + a, y + b) \in G$ for $(a, b) \in \mathbb{Z}^2$.
(2) Because more than one class is used to find an appropriate difference operator, there must be a clear *structural relation between distinct classes*.
(3) It must be straightforward to verify *class membership*, that is, given f and G, verifying whether f is in G must be simple.
(4) It must be straightforward to verify *fitting class membership*, that is, given f and G and a small real number $\epsilon > 0$ one must be able to verify whether there exists a function $\tilde{g} \in G$ such that $\lfloor f - \tilde{g} \rfloor < \epsilon$, without actually calculating $\tilde{g}$.
(5) There must be an *appropriate difference operator* Q that can be used to approximate the differential operator L for the entire class G, that is, $Q\tilde{g} = L\tilde{g}$ for every $\tilde{g} \in G$.
(6) When more than one operator Q satisfies the constraint $Q\tilde{g} = L\tilde{g}$, then there must be an additional criterion that can be used to choose an optimal operator.

Although the above criteria seem to be quite strict, classes of fitting functions exist that meet all the requirements. We will define a class of fitting functions $\tilde{g}$ as the set of functions $\tilde{g}$ that satisfy a given system of difference equations $P_i\tilde{g} = 0$.

We first note that such a system of difference equations is not unique. In fact, for any operator P in $I = \langle P_1, P_2, \ldots \rangle$, we have $P\tilde{g} = 0$. For example, $P_1P_2\tilde{g} = 0$, $(P_1 + P_2)\tilde{g} = 0$, and $(\sum S_iP_i)\tilde{g} = 0$ for arbitrary difference operators $S_i \in \mathbb{R}[\sigma]$. There is thus more than one way to write down the difference equations that define a class. Therefore, it makes sense to view the entire ideal $I = \langle P_1, P_2, \ldots \rangle$ as the defining object for a class of fitting functions. We now show that classes of fitting functions defined by ideals of difference equations satisfy all the necessary requirements.

2.1. Shift-invariance

Suppose that $\tilde{g}$ satisfies the difference equations $P_i\tilde{g}(x, y) = 0$ of a given class. Then $\tilde{g}$ also satisfies $\sigma_x P_i\tilde{g}(x, y) = 0$. Since the P_i do not depend on x and y, it follows that $P_i\sigma_x\tilde{g}(x, y) = 0$, and therefore $P_i\tilde{g}(x + 1, y) = 0$. Similarly, we have $P_i\tilde{g}(x, y + 1) = 0$, and it follows that the class is shift-invariant. Note that shift-invariance does not hold when the operators P_i depend on x or y.

2.2. Structural relations between classes

Clearly, for two ideals I_1 and I_2 such that $I_1 \subseteq I_2$, we have $G_2 \subseteq G_1$. Furthermore, if $I_1 = \langle P_1, \ldots, P_k \rangle$ defines G_1, and $I_2 = \langle P_{k+1}, \ldots, P_s \rangle$ defines G_2, then the ideal $I = \langle P_1, \ldots, P_k, P_{k+1}, \ldots, P_s \rangle$ defines $G_1 \cap G_2$. The relations that exist between distinct classes, however, extend much further.

Groebner bases. An ideal I of difference operators may be generated by a possibly infinite set of operators P_i. Hilbert's Basis Theorem for polynomial ideals states that any ideal of polynomials in the ring $\mathbb{R}[\sigma_x, \sigma_y]$ can always be generated by a finite set of polynomials. Even if the defining system has infinitely many difference equations, these can all be obtained by multiplying, adding and translating a finite set of equations. The best way to find a small set of defining polynomials is to compute a Groebner basis for the ideal [11]. A Groebner basis can be computed for any ideal by imposing a monomial ordering relation on the shift operators σ_x, σ_y [11], for example lexicographic order (or lex order). In this paper, the choice of the ordering relation will not be important; the only consequence will be that the Groebner bases are not necessarily symmetrical in the shift operators σ_x, σ_y. Note that a Groebner basis can be computed either for polynomials in σ_x, σ_y, or for polynomials in the forward differences Δ_x, Δ_y, with $\Delta_x = \sigma_x - 1$ and $\Delta_y = \sigma_y - 1$. The result will be the same after the substitutions $\Delta_x^j = (\sigma_x - 1)^j$ and $\Delta_y^j = (\sigma_y - 1)^j$. In terms of templates, for a given class of fitting functions, a Groebner basis completely characterizes all different templates that will recognize a function of that class [9].

Reduced Groebner bases. The most basic property of a Groebner basis is that it can be used in a division algorithm to determine whether a given polynomial belongs to an ideal. A polynomial belongs to an ideal when the (unique) remainder is zero after division by the polynomials in the Groebner basis. Furthermore, for a given order, most computer algebra systems compute a reduced Groebner basis, which is unique and minimal. With reduced Groebner bases we can answer several questions regarding classes of fitting functions and the difference equations they satisfy.

Identical classes. Two systems of difference equations correspond to the same class of fitting functions, when the reduced Groebner bases are the same.

Class subset relations. How can we determine whether a class is a subset of another class? Clearly, $G_1 \subset G_2$ if and only if $I_1 \supset I_2$. Therefore it suffices to verify whether one ideal is a subset of another ideal. With reduced Groebner bases this is straightforward. Suppose that we have the following ideals: $I_1 = \langle \Delta_x - \Delta_y, \Delta_y^2 \rangle$ and $I_2 = \langle \Delta_x^2, \Delta_x\Delta_y, \Delta_y^2 \rangle$. The reduced Groebner basis of the ideal generated by $\langle \Delta_x - \Delta_y, \Delta_y^2, \Delta_x^2, \Delta_x\Delta_y, \Delta_y^2 \rangle$ is equal to $\langle \Delta_x - \Delta_y, \Delta_y^2 \rangle$, which is the same as the reduced basis for I_1. Furthermore, because the reduced bases for I_1 and I_2 are distinct, it follows that $I_1 \supset I_2$, and therefore $G_1 \subset G_2$.

One way to define a hierarchy of fitting classes is to start with an ideal of operators, and add one operator at a time to extend the ascending chain of ideals: $I_1 \subset I_2 \subset I_3 \subset \cdots$, implying $G_1 \supset G_2 \supset G_3 \supset \cdots$. The ascending chain condition states that this chain will stabilize, that is, $I_N = I_{N+1} = \ldots$ after a while. Ascending chains of classes will be used later on.

Non-zero solutions. We can verify whether a system of difference equations has a solution other than the zero solution. For example, a Groebner basis for the ideal generated by $I = \langle \sigma_x + \sigma_y, \sigma_x - 1, \sigma_y - 1 \rangle$ is $\langle 1 \rangle$. This means that the ideal $I = \langle 1 \rangle$ contains all possible polynomials in σ_x, σ_y, and that the zero function is the only solution of the difference equations $(\sigma_x + \sigma_y)\tilde{g} = 0$, $(\sigma_x - 1)\tilde{g} = 0$, $(\sigma_y - 1)\tilde{g} = 0$.

2.3. Class membership

Since an ideal of operators is used to define a class of fitting functions, in principle, we should verify an infinite number of difference equations. Fortunately, class membership can always be determined by verifying a small number of difference equations. With a reduced Groebner basis any polynomial in the ideal can be decomposed into polynomials in the Groebner basis, using a division algorithm. Thus a reduced Groebner basis provides the minimum number of difference equations $P_1 f = 0, \ldots, P_k f = 0$, such that if f satisfies these equations, then it satisfies any equation $Pf = 0$, with P in the ideal.

Alternatively, we can start with a class of fitting functions that is shift-invariant, and write down a sufficient number of difference equations that are satisfied by any member of the class. The reduced Groebner basis then yields a minimal basis that defines the fitting class.

Example. The functions of the form

$$\tilde{g}(x, y) = \alpha_1(x + y)^2 + \alpha_2(x + y) + \alpha_3 \tag{3}$$

form a shift-invariant class, in which each function satisfies the following difference equations: $\Delta_x^3\tilde{g} = 0$, $\Delta_y^3\tilde{g} = 0$, $(\Delta_x^2 - \Delta_y^2)\tilde{g} = 0$, $(\Delta_x - \Delta_y)\tilde{g} = 0$. Here we have included a sufficient number of difference equations, so that all solutions are of the required form (3). How a sufficient number of generating equations can be written down is explained in Section 2.4. The reduced Groebner basis for the ideal $I = \langle \Delta_x^3, \Delta_y^3, \Delta_x^2 - \Delta_y^2, \Delta_x - \Delta_y \rangle$ is given by $\langle \Delta_x - \Delta_y, \Delta_y^3 \rangle$.

2.4. Fitting class membership

Given $\epsilon > 0$, and a neighborhood D, we will say that f belongs to the fitting class of G_ϵ if there is a $\tilde{g}$ in G, such that

$$|f - \tilde{g}| < \epsilon, \tag{4}$$

in D. One can prove the following result [9].

Proposition 1. *Let f be a function, and G a class of fitting functions defined by an ideal of difference operators I. Then $f \in G_\epsilon$ if and only if*

$$|Pf| < \epsilon \sum |p_j| \tag{5}$$

for every operator P in the ideal I.

In fact, suppose that $|f - \tilde{g}| < \epsilon$, then it follows immediately that for any operator $P \in I$, $|Pf - P\tilde{g}| < \epsilon \sum |p_j|$, and therefore $|Pf| < \epsilon \sum |p_j|$. That is, any operator $P \in I$ can be used to eliminate the fitting function $\tilde{g}$ from the inequality (4). That the converse is also true is shown in [9].

Thus the introduction of the difference equations $P_i\tilde{g} = 0$ has an important consequence: by eliminating the explicit occurrence of $\tilde{g}$, the explicit computation of a fitting function is replaced by the use of the operators P_i. Instead of actually fitting $\tilde{g}$ to f, it is sufficient to verify whether (5) holds for every difference operator in I, i.e. whether f belongs to the class G_ϵ. So we do not bother which fitting function in G would actually yield the closest fit.

According to Proposition 1, $|Pf| < \epsilon \sum |p_j|$ must be verified for an infinite number of operators P, a process which can be simplified. Fitting class membership can be established by verifying only a finite number of inequalities, provided the solution space of the difference equations is a finite linear vector space. Assume that the solution set of the partial difference equations $P_1 g = 0, \ldots, P_n g = 0$ can be written as a linear vector space with $g_1, \ldots, g_l$ as a basis:

$$\alpha_1 g_1 + \cdots + \alpha_l g_l. \tag{6}$$

Let K_D be the set of all difference operators P_i of the form

$$\begin{vmatrix} g_1(x_1) & \ldots & g_l(x_1) & \sigma^{x_1} \\ \ldots & & & \\ g_1(x_{l+1}) & \ldots & g_l(x_{l+1}) & \sigma^{x_{l+1}} \end{vmatrix} \tag{7}$$

with the points $x_j \in D$. The operators of K_D are written as determinantal expressions of the coefficients $g_j(x_j)$ and the shift operators σ^{x_j}. Since D is finite, K_D is finite. The operators P in K_D are all operators for which Pf is well defined. When D is rectangular, there is a unique point x_0 in D with smallest coordinates. Let $\sigma^{-x_0}K_D$ denote the polynomials of K_D multiplied by σ^{-x_0}. The polynomials $\sigma^{-x_0}K_D$ all have non-negative exponents, but some of the exponents will be zero. If D is chosen not too small then the polynomials in $\sigma^{-x_0}K_D$ will generate the entire ideal $I = \langle P_1, \ldots, P_n \rangle$. In practice, for two-dimensional operators, it is usually sufficient that D has size at least $(a+2) \times (b+2)$ when the leading monomial of the fitting functions is $x^a y^b$, or smaller. In that case one can show that Proposition 1 can be extended, so that for f to be in G_ϵ it is sufficient to verify (5) for all P in the finite set K_D [9,12].

When the operators in K_D form a generating subset for the ideal that defines the class G of fitting functions of the form (6), K_D can also be used as a starting point to compute a reduced Groebner basis for I. In fact, a small sample of K_D is sufficient.

2.5. Appropriate difference operators

Given a class of fitting functions G, a difference operator Q must be found such that $Q\tilde{g} = L\tilde{g}$ for all $\tilde{g}$ in the fitting class G. Can such a difference operator always be found? A Taylor expansion yields the following relation between the shift operator σ_x and partial derivatives:

$$\sigma_x = 1 + \frac{\partial}{\partial x} + \frac{1}{2!}\frac{\partial^2}{\partial x^2} + \frac{1}{3!}\frac{\partial^3}{\partial x^3} + \cdots. \tag{8}$$

It follows that

$$\Delta_x = \frac{\partial}{\partial x} + \frac{1}{2!}\frac{\partial^2}{\partial x^2} + \frac{1}{3!}\frac{\partial^3}{\partial x^3} + \cdots. \tag{9}$$

Thus, given a class of polynomials P with bounded degree and a difference operator Q, there is a differential operator L such that $Q\tilde{g} = L\tilde{g}$. Likewise, one can prove the following symbolic expansion:

$$\frac{\partial}{\partial x} = \Delta_x - \frac{\Delta_x^2}{2} + \frac{\Delta_x^3}{3} - \frac{\Delta_x^4}{4} + \cdots. \tag{10}$$

This shows that for a given differential operator L we can always find a difference operator Q such that $Q\tilde{g} = L\tilde{g}$.

Example. According to (10) the Laplacian $L = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}$ can be replaced by

$$\left(\Delta_x - \frac{\Delta_x^2}{2} + \frac{\Delta_x^3}{3} + \cdots\right)^2 + \left(\Delta_y - \frac{\Delta_y^2}{2} + \frac{\Delta_y^3}{3} + \cdots\right)^2. \tag{11}$$

For polynomial functions $\tilde{g}(x, y)$ of degree at most 2, the nth order difference Δ^n is zero for $n \geq 3$. After expansion, and removal of all differences of order higher than two, it follows that $L\tilde{g}(x, y) = (\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2})\tilde{g}(x, y) = (\Delta_x^2 + \Delta_y^2)\tilde{g}(x, y) = ((\sigma_x - 1)^2 + (\sigma_y - 1)^2)\tilde{g}(x, y)$ for all polynomials of degree 2 or smaller.

2.6. *Optimal operators*

As mentioned before, the choice of Q is not unique. Since $(Q + P)\tilde{g} = Q\tilde{g}$ for any operator P in the ideal I, there are infinitely many choices for Q. There are, however, no other possibilities than those provided by the ideal I. Every operator R satisfying $R\tilde{g} = Q\tilde{g}$ can be written as $R = Q + P$.

Since $R\tilde{g} = L\tilde{g}$ and $|f - \tilde{g}| < \epsilon$, it follows that

$$|Rf - L\tilde{g}| < \epsilon \sum |r_j|, \tag{12}$$

where the r_j are the coefficients of R. Among all operators of the form $R = P+Q$, we will therefore look for the best candidate. More precisely, we look for an operator of the form

$$R = Q + \sum S_i P_i, \tag{13}$$

where $S(\sigma_x)$ is an arbitrary difference operator such that $\sum |r_j|$ is as small as possible. This gives a systematic method for computing the optimal difference operator for a particular fitting function class.

Example. For the class of functions of the form

$$\alpha_1(x + y)^2 + \alpha_2(x + y) + \alpha_3 \tag{14}$$

the reduced Groebner basis is $\langle \Delta_x - \Delta_y, \Delta_y^3 \rangle$. The Laplacian $(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2})$ can be replaced by any difference operator of the form

$$R(\sigma_x, \sigma_y) = (\sigma_x - 1)^2 + (\sigma_y - 1)^2 + S_1(\sigma_x - \sigma_y) + S_2(\sigma_y - 1)^3 \tag{15}$$

where S_1, S_2 are arbitrary operator polynomials in σ_x, σ_y. We must choose S_1, S_2 such that the sum of the absolute values of the coefficients $R(\sigma_x, \sigma_y) = \sum |r_j|$ becomes minimal. In this work, this kind of minimization problem has been solved by exhaustive computer search. An optimal operator can be found by solving a large set of linear programming problems. Although the expression $\sum |r_j|$ is nonlinear, the optimum can be found by solving 2^n small linear programming problems, with n the number of terms with an absolute value. In each problem we make an assumption for each absolute value about the sign of the term, and replace it by a linear term.

If we restrict the search space to polynomials of the form $S_1 = a_1\sigma_x^2 + a_2\sigma_y^2 + a_3\sigma_x\sigma_y + a_4\sigma_x + a_5\sigma_y + a_6$, and $S_2 = a_7$, we find that the best operator is $R(\sigma_x, \sigma_y) = 1 - \sigma_x - \sigma_y^2 + \sigma_y^3$. This corresponds to the template

$$\begin{array}{|c|c} \hline 1 & \\ \hline -1 & \\ \hline 0 & \\ \hline 1 & \multicolumn{1}{|c|}{-1} \\ \hline \end{array} \tag{16}$$

This template is not symmetrical. Because the form of the fitting functions as well as the Laplacian is symmetrical in x and y, and because the fitting class is shift-invariant, we may add appropriate reflections of the template, and divide the result

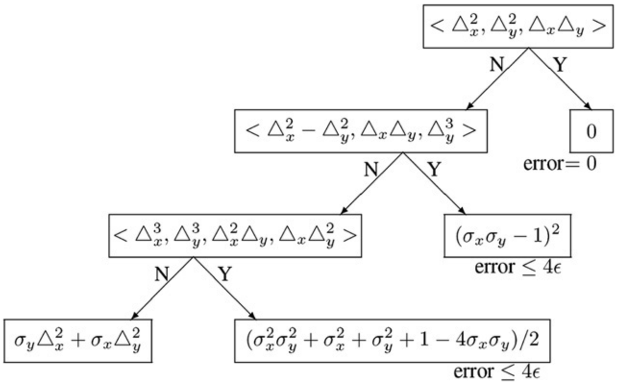

Fig. 1. A decision chain for the Laplacian.

by the number of templates that have been added. The result is the following symmetrical template:

$$\frac{1}{4} * \begin{array}{|c|c|c|c|c|c|c|} \cline{4-4} \multicolumn{3}{c|}{} & 1 & \multicolumn{3}{c}{} \\ \cline{4-4} \multicolumn{3}{c|}{} & -1 & \multicolumn{3}{c}{} \\ \cline{4-4} \multicolumn{3}{c|}{} & -1 & \multicolumn{3}{c}{} \\ \hline 1 & -1 & -1 & 4 & -1 & -1 & 1 \\ \hline \multicolumn{3}{c|}{} & -1 & \multicolumn{3}{c}{} \\ \cline{4-4} \multicolumn{3}{c|}{} & -1 & \multicolumn{3}{c}{} \\ \cline{4-4} \multicolumn{3}{c|}{} & 1 & \multicolumn{3}{c}{} \\ \cline{4-4} \end{array} \tag{17}$$

The sum of the absolute values of the template coefficients is 4. The sum of the absolute values of $(\sigma_x - 1)^2 + (\sigma_y - 1)^2$ is 8, which means that when we compute the Laplacian using (17), the error will be twice as small as when we use the standard template to compute the Laplacian.

3. A decision chain for the Laplacian operator

In this section we apply the above theory in the design of appropriate difference operators for the widely used Laplacian $\partial^2/\partial x^2 + \partial^2/\partial y^2$. To determine the correct fitting function class in each image point, we use the decision chain shown in Fig. 1, which has two different kinds of nodes.

The first kind of node shows the Groebner basis for the difference operators that define a function class. The second kind of node shows for each class the difference operator that replaces the differential. When an image is being processed, at each pixel, a small set of inequalities (typically less than 10) is verified for some small rectangular neighborhood (typically 5×5). These inequalities are of the form $|Pf| < \epsilon \sum |p_j|$, where P is a polynomial from the operator ideal that defines the function class. When all the inequalities are satisfied, the appropriate function class has been localized, and the optimal difference operator computes the Laplacian.

The motivation for using a decision chain is based on ascending chains of fitting classes. Suppose that we have a chain of fitting classes $G_1 \subset G_2 \subset G_3 \subset \cdots$. Then we can set up a decision chain where we first verify whether a function f belongs to $G_{1;\epsilon}$. If $f \in G_{1;\epsilon}$ we apply the optimal operator for $G_{1;\epsilon}$. If $f \notin G_{1;\epsilon}$ we proceed by verifying whether $f \in G_{2;\epsilon}$, and so on. The meaning of this decision chain is implied by the reverse chain of ideals $I_1 \supset I_2 \supset I_3 \supset \cdots$. Since $I_n \supset I_{n+1}$, when computing optimal operators of the form $R_j = Q + P_j$, with $P_j \in I_j$, the search space for R_n is a superset of the search space for R_{n+1}, and therefore the error made by the best operator for the class G_n will be smaller than, or at most equal to, the error for the class G_{n+1}.

The decision chain of Fig. 1 uses three fitting function classes, for which $G_1 \subset G_2 \subset G_3$:

$$\begin{aligned} &G_1 : \alpha_1 x + \alpha_2 y + \alpha_3, \\ &G_2 : \alpha_1(x^2 + y^2) + \alpha_2 x + \alpha_3 y + \alpha_4, \\ &G_3 : \alpha_1 x^2 + \alpha_2 y^2 + \alpha_3 xy + \alpha_4 x + \alpha_5 y + \alpha_6. \end{aligned} \tag{18}$$

The Groebner bases and difference operators are obtained using the techniques of the previous section. A set of feature detection templates is computed for each class of fitting functions, using the polynomials (7). We illustrate this for the

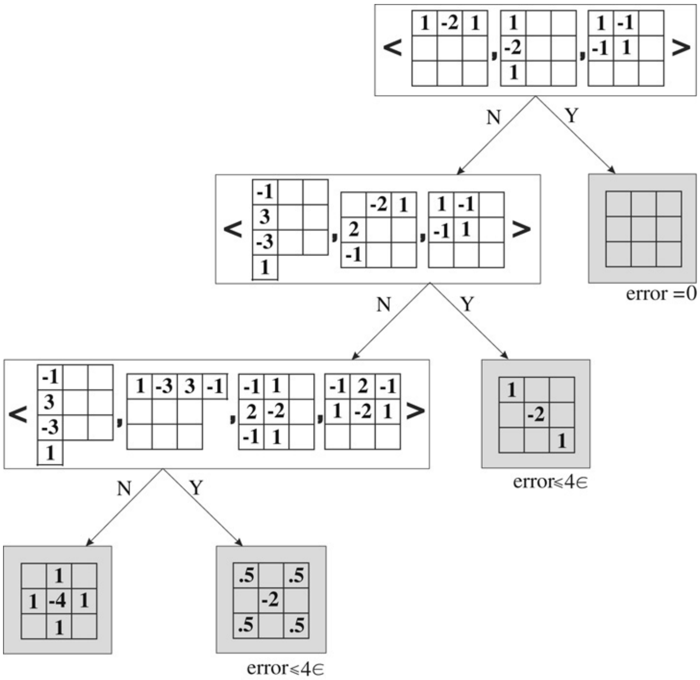

Fig. 2. A decision chain for the Laplacian represented by templates.

function class in the second node of the decision chain: $\alpha_1(x^2+y^2)+\alpha_2 x+\alpha_3 y+\alpha_4$. In this case, (7) takes the form

$$\begin{vmatrix} 1 & x_1 & y_1 & x_1^2+y_1^2 & \sigma_x^{x_1} & \sigma_y^{y_1} \\ \cdots & & & & & \\ 1 & x_5 & y_5 & x_5^2+y_5^2 & \sigma_x^{x_5} & \sigma_y^{y_5} \end{vmatrix} \tag{19}$$

For example, for $\{(x_1,y_1),\dots,(x_5,y_5)\}=\{(0,0),(1,1),(2,0),(3,1),(4,0)\}$, the above determinant is equal to $1-2\sigma_x\sigma_y+2\sigma_x^3\sigma_y-\sigma_x^4$ which corresponds to the template

$$\begin{array}{|c|c|c|c|c|} \hline 1 & & \square & & -1 \\ \hline & -2 & & 2 & \\ \hline \end{array} \tag{20}$$

Thus the polynomials (19) form a generating subset for the ideal I_D of G_2. In fact, a small sample of these polynomials is sufficient to compute the Groebner basis $\langle(\sigma_x-1)^2-(\sigma_y-1)^2,(\sigma_x-1)(\sigma_y-1),(\sigma_y-1)^3\rangle$, which can also be written as $\langle\Delta_x^2-\Delta_y^2,\Delta_x\Delta_y,\Delta_y^3\rangle$.

The corresponding templates are

$$\begin{array}{|c|c|c|} \hline & -2 & 1 \\ \hline 2 & & \\ \hline -1 & & \\ \hline \end{array}\ ,\quad \begin{array}{|c|c|} \hline 1 & -1 \\ \hline -1 & 1 \\ \hline \end{array}\ ,\quad \begin{array}{|c|} \hline -1 \\ \hline 3 \\ \hline -3 \\ \hline 1 \\ \hline \end{array}\ . \tag{21}$$

For each of the fitting function classes in Fig. 1, the Groebner basis completely characterizes the templates needed to classify the function. Fig. 2 shows the same decision chain as Fig. 1, but now with templates representing Groebner bases, instead of polynomials. Table 1 gives a larger list of function classes and their Groebner bases.

Optimal difference operators. Once the template bases for the polynomial ideals are known, the next step is to determine an optimal difference operator for each class of fitting functions. As an example we describe the computation of the difference operator for quadratic functions of the form $\alpha_1x^2+\alpha_2y^2+\alpha_3xy+\alpha_4x+\alpha_5y+\alpha_6$. We must choose Q so that the requirement $L\tilde{g}=Q\tilde{g}$ is satisfied, that is

$$\left(\frac{\partial^2}{\partial x^2}+\frac{\partial^2}{\partial y^2}\right)\tilde{g}=(\Delta_x^2+\Delta_y^2)\tilde{g}=2\alpha_1+2\alpha_2. \tag{22}$$

Table 1
Function classes and corresponding Groebner bases with lexicographic ordering $\Delta_x > \Delta_y$

α_1	$\langle \Delta_x, \Delta_y \rangle$
$\alpha_1 x + \alpha_2$	$\langle \Delta_x^2, \Delta_y \rangle$
$\alpha_1(x+y) + \alpha_2$	$\langle \Delta_x - \Delta_y, \Delta_y^2 \rangle$
$\alpha_1 x + \alpha_2 y + \alpha_3$	$\langle \Delta_x^2, \Delta_x \Delta_y, \Delta_y^2 \rangle$
$\alpha_1 xy + \alpha_2 x + \alpha_3 y + \alpha_4$	$\langle \Delta_x^2, \Delta_y^2 \rangle$
$\alpha_1(x+y)^2 + \alpha_2(x+y) + \alpha_3$	$\langle \Delta_x - \Delta_y, \Delta_y^3 \rangle$
$\alpha_1(x^2+y^2) + \alpha_2 x + \alpha_3 y + \alpha_4$	$\langle \Delta_x^2 - \Delta_y^2, \Delta_x \Delta_y, \Delta_y^3 \rangle$
$\alpha_1(x+y)^2 + \alpha_2 x + \alpha_3 y + \alpha_4$	$\langle \Delta_x^2 - \Delta_y^2, \Delta_y(\Delta_x - \Delta_y), \Delta_y^3 \rangle$
$\alpha_1 x^2 + \alpha_2 y^2 + \alpha_3 x + \alpha_4 y + \alpha_5$	$\langle \Delta_x^3, \Delta_x \Delta_y, \Delta_y^3 \rangle$
$\alpha_1 x^2 + \alpha_2 y^2 + \alpha_3 xy + \alpha_4 x + \alpha_5 y + \alpha_6$	$\langle \Delta_x^3, \Delta_x^2 \Delta_y, \Delta_x \Delta_y^2, \Delta_y^3 \rangle$

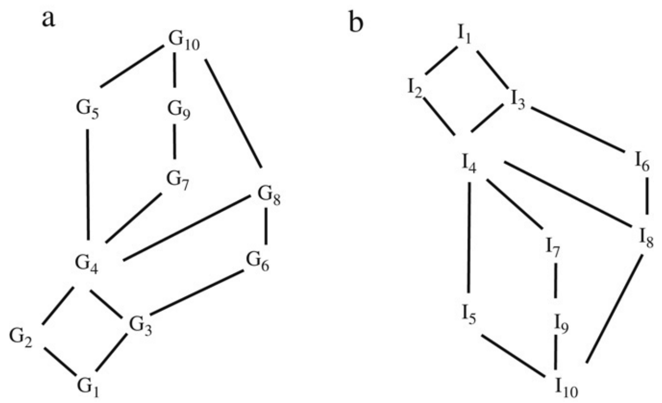

Fig. 3. Partial order of fitting function classes: The subset relations that exist between the classes induce a partial order.

For this function class, any differential operator of the form

$$R(\sigma_x, \sigma_y) = \Delta_x^2 + \Delta_y^2 + \Delta_x^3 S_1 + \Delta_y^3 S_2 + \Delta_x^2 \Delta_y S_3 + \Delta_x \Delta_y^2 S_4 \tag{23}$$

in which S_i are arbitrary polynomials in the shift operators, yields the exact value for the Laplacian. For optimal results, one must choose the operator of the form (23) for which $\sum |r_j|$ is as small as possible. By solving the minimization problem we obtain the symmetric operator

$$\frac{\sigma_x^2 \sigma_y^2 + \sigma_x^2 + \sigma_y^2 + 1 - 4\sigma_x \sigma_y}{2} \tag{24}$$

as the optimal choice for this function class. The difference operators for the other function classes are obtained in a similar way, and their templates are shown in Fig. 2.

The obtained difference operators lead to some interesting conclusions. Without feature detection, we would always choose the classical discrete equivalent of the Laplacian as shown in the lower left leaf of the decision chain. When the digitized function is locally linear or close to linear, the best possible difference operator is the zero operator, yielding an error equal to zero, because the Laplacian of a linear function vanishes. Since difference operators for linear functions do not have to compensate for quadratic terms, they perform better than the difference operators derived for quadratic functions. For both quadratic functions in Fig. 1, there is a difference operator which yields a maximal error of 4ϵ on the computed value of the Laplacian. For quadratic functions with circular symmetry (the second level in the decision chain), it is sufficient to compute the second-order difference in a diagonal direction. Finally, for quadratic functions the best difference operator has a template equal to the classical discrete Laplacian operator rotated over 45° and divided by 2. If none of the feature detection tests succeed, we use the classical discrete Laplacian operator. In fact, the differences between the operators appear to be quite small, but the examples in Section 4 will show that considerable improvements are obtained by using the decision chain.

Posets of fitting classes. In general, a set of fitting classes does not form a linear, ascending chain, but it forms a partially ordered set or poset, with subset inclusion as the ordering relation. Fig. 3a shows the Hasse-diagram for the subset relations between the classes of Table 1. By turning Fig. 3a upside-down, we obtain in Fig. 3b the subset order for the ideals I_j.

The ideals form a join-semilattice, that is for each pair of ideals I_j, I_k, the join $I_j \vee I_k$ is also part of the semilattice, where the join is defined as $I_j \vee I_k = I_j \cup I_k$. For example, $I_5 \cup I_6 = I_3$, which can easily be verified by computing a reduced Groebner basis for $\langle \Delta_x^2, \Delta_y^2, \Delta_x - \Delta_y, \Delta_y^3 \rangle$. The ideals do not form a meet-lattice, and therefore not a lattice. A Groebner basis for the intersection of two ideals I_j, I_k with Groebner bases $\langle P_1, \ldots, P_n \rangle$ and $\langle Q_1, \ldots, Q_m \rangle$, respectively, can be found by taking a

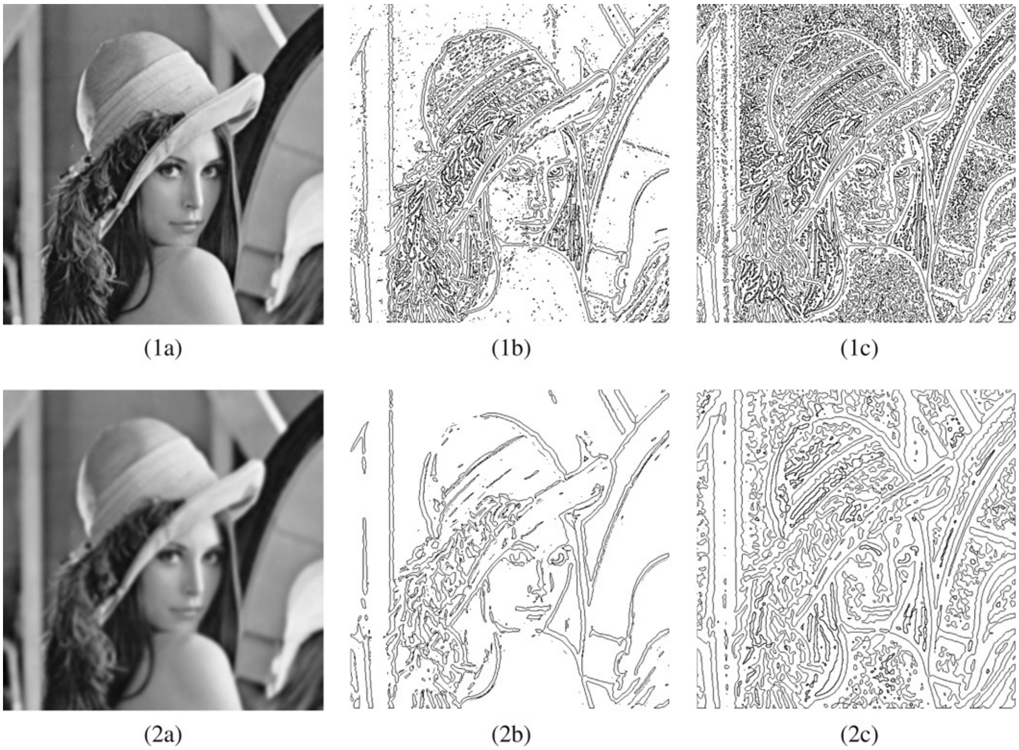

Fig. 4. LoG edge detection: In column (a) we see the Gaussian smoothed images on different scales, respectively for standard deviation $\sqrt{2}$ and $2\sqrt{2}$. Column (b) shows the result of edge detection after application of the decision chain to compute the Laplacian with $\epsilon = 1/2$ in the examples shown. Column (c) shows the result of edge detection after application of the classical Laplacian kernel.

Groebner basis for the ideal $\langle tP_1, \dots, tP_n, (1-t)Q_1, \dots, (1-t)Q_m \rangle$, relative to any lexicographic ordering which puts t in front. The polynomials in the basis that do not contain t form a Groebner basis for $I_j \cap I_k$ [11]. In this case, the reduced Groebner basis for $I_7 \cap I_8$ is equal to $\langle \Delta_x^2 - \Delta_y^2, \Delta_x\Delta_y^2, \Delta_y^3 \rangle$. This ideal and its corresponding function class, with functions of the form $\alpha_1(x^2+y^2) + \alpha_2 xy + \alpha_3 x + \alpha_4 y + \alpha_5$, are not shown in Table 1.

Fig. 3 suggests the possibility to compute the Laplacian with an elaborate poset of function classes, instead of the simple linear decision chain of Fig. 1. In a poset, however, the choice of an optimal difference operator may not be unique. A function f can belong to both $G_{2;\epsilon}$ and $G_{3;\epsilon}$, but not to the class $G_{1;\epsilon}$. In this case, additional criteria are needed to select the difference operator either from class $G_{2;\epsilon}$ or from $G_{3;\epsilon}$.

4. Laplacian of Gaussian edge detection

We illustrate the use of the decision chain, presented in the previous section to compute edges. Edge pixels can be found as the zero crossings of a Laplacian, after Gaussian smoothing (LoG). A LoG is computed by first convolving the image with a Gaussian kernel of a certain width and then passing a Laplacian filter kernel over the Gaussian smoothed image. The results of the decision chain with the optimal difference operators are compared to those obtained with the classical non-optimized version of the kernel. In a decision chain, the local characteristics of the digitized image surface are determined by subsequently verifying a set of inequalities for each class of fitting functions. Once the correct fitting function has been determined, the corresponding difference operator is applied. For each image in the experiment, we constructed a scale space for different widths, i.e. different values for the standard deviation, of the Gaussian kernel, so that we can compare both methods for increasing levels of smoothing. The results for the edge detection by both methods are shown in Fig. 4.

Are fitting function classes useful? For each function class, a set of inequalities or templates is created to detect the local features of the digitized image. When the image surface locally fulfills the conditions posed by a limited subset of inequalities of the form (5), with $\epsilon = 1/2$ to account for digitization errors, we consider the fitting function a good approximation for the surface. Fig. 5 shows the usefulness of the three function classes in the decision chain. The function class to which a pixel belongs is indicated for each pixel by a gray value: the higher the intensity, the lower the class occurs in the decision chain. Where none of the continuous fitting functions approximated the image surface well, e.g. at sharp or discontinuous edges, the pixels are indicated as white. Then the classical version of the Laplacian kernel is applied. We cannot define or predict the error on the computation of the Laplacian in such points. Since all gray values appear in Fig. 5, all the nodes of

Fig. 5. The fitting function class in each image point is indicated: The higher (whiter) the point's intensity, the lower the node of the fitting function appears in the decision chain of Fig. 2. When no correct approximation is found, points are colored white. The image was convolved first with a Gaussian of standard deviation $\sqrt{2}$.

the decision chain were used repeatedly. Large parts in Fig. 5 are black, which means that the image is almost linear in these parts, and that the best approximation to the Laplacian is the zero value.

How many inequalities must be verified when traversing the decision chain? The local image features are detected in a neighborhood D by verifying a set of inequalities of the form (5). For each node of the decision chain, we generate a set K_D of operators P as in (7). The total amount of inequalities, i.e. the size of the set K_D, is equal to $\binom{d}{l}$, with d the number of points in the neighborhood D and l the number of basis vectors in the solution set of $P_i\tilde{g} = 0$ as in (6). Verifying a set of that size at each node yields a rather large computational complexity for the decision chain. The complexity is reduced by verifying only a small amount of inequalities at each node. The results shown in Fig. 4 are obtained with a subset of seven feature templates per node. Simulations show that the classification into function classes does not significantly improve when more than seven inequalities are used, only for a small number of pixels are the results different. Although a random selection yields good results, there is an optimal way to select a small number of inequalities as discussed in [9,13].

The computational complexity of the decision chain can be further reduced by reusing feature templates in more than one neighborhood D. The computation of the templates is dependent on the size of D, but most of the polynomial's coefficients are zero, as can be seen in (20). So the majority of the results for the verification of inequalities can be reused in adjacent neighborhoods. To avoid systematic errors, introduced by using the same subset of templates over and over for all points, we add a few random subsets in each neighborhood to complement the reused results.

The Laplacian computed with the proposed decision chain has been compared with the classical Laplacian kernel. Several images were analyzed in a scale space of Gaussians with increasing standard deviation. Fig. 4 shows the results for Gaussians with standard deviation $\sqrt{2}$ and $2\sqrt{2}$. The results for the LoG edge detection are considerably better when the Laplacian is computed with the decision chain. The most significant edges are detected and the edges of important details are preserved, as the images in column (b) show. If we compare this to the result in column (c), we notice an abundance of edges in image regions considered homogeneous. Even on a higher scale, i.e. for even smoother images, the computation of the Laplacian with the classical kernel does not yield better results. First, the error on the localization of edges increases for higher levels of smoothing. Second, the edges of finer (and even coarser) details disappear on higher scales while edges are still detected in (noisy) homogeneous regions. Both problems are avoided with the decision chain. Edges of details are already distinguished on finer scales. If linear fitting functions can locally approximate the image surface in homogeneous regions, the zero operator is used for the Laplacian so that zero crossings do not occur in these regions.

5. Conclusion

We present a mathematical framework from which both feature detection and difference operators arise in a natural way. By detecting local image features, we avoid the necessity of actually approximating the digitized image surface by fitting functions. For each function class, we define the appropriate difference operator which yields a minimal computational error when approximating the value we would have obtained by the differential operator. Some issues should be further examined. When a poset of function classes is used, additional criteria are needed to decide which function class has to be used when there is a draw between classes. Also the choice of neighborhood size and shape has not been discussed in this paper. Would we prefer an approximation in a large neighborhood to an approximation in a smaller neighborhood, when the latter is slightly better? Also here additional criteria are necessary. We conclude that the quality of the edge pixels detected by the LoG improves when the Laplacian difference operator is adapted to the local image features. Clearly, a practical application like the computation of straight lines and corners of a building as in Fig. 6a, is considerably easier when given the LoG edge pixels of Fig. 6b (decision chain) as opposed to the information obtained by the default kernel, shown in Fig. 6c.

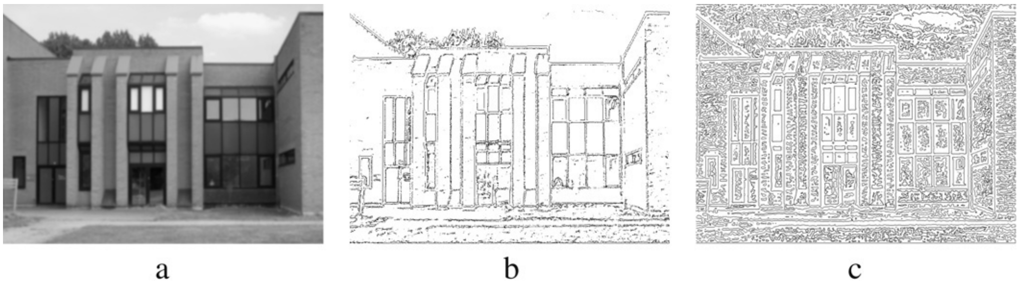

Fig. 6. LoG edge detection: Image (a) shows a Gaussian smoothed image with standard deviation $2\sqrt{2}$. Image (b) shows the result of edge detection after application of the decision chain to compute the Laplacian. Image (c) shows the result of edge detection after application of the classical Laplacian kernel.

References

[1] R. Haralick, Digital step edges from zero crossing of second directional derivatives, IEEE Trans. Pattern Anal. Mach. Intell. 6 (1984) 58–68.
[2] D. Langridge, Detection of discontinuities in the first derivatives of surfaces, Comput. Vis. Graph. Image Process. 27 (1984) 291–308.
[3] M. Fleck, Multiple widths yield reliable finite differences, IEEE Trans. Pattern Anal. Mach. Intell. 14 (1992) 412–429.
[4] E. Karabassis, M.E. Spetsakis, An analysis of image interpolation, differentiation, and reduction using local polynomial fits, CVGIP: Graph. Models Image Process. 57 (1995) 183–196.
[5] J.-O. Lachaud, A. Vialard, F. de Vieilleville, Analysis and comparative evaluation of discrete tangent estimators, in: Proc. DGCI'05, in: LNCS, vol. 3429, 2005, pp. 240–251.
[6] T. Lindeberg, Discrete derivative approximations with scale-space properties: A basis for low-level feature extraction, J. Math. Imaging Vis. 3 (1993) 349–376.
[7] S. Gunn, On the discrete representation of the Laplacian of Gaussian, Pattern Recognit. 32 (1999) 1463–1472.
[8] D. Demigny, T. Kamlé, A discrete expression of Canny's criteria for step edge detector performances evaluation, IEEE Trans. Pattern Anal. Mach. Intell. 19 (1997) 1199–1211.
[9] P. Veelaert, Local feature detection for digital surfaces, in: Proceedings of the SPIE Conference on Vision geometry V, SPIE 2826 (1996) 34–45.
[10] K. Teelen, P. Veelaert, Improving difference operators by local feature detectors, in: Lecture Notes in Computer Science: Proc. Discrete Geometry for Computer Imagery, vol. 4245, Springer, Szeged, 2006, pp. 391–402.
[11] D. Cox, J. Little, D. O'Shea, Ideals, Varieties and Algorithms: An Introduction to Computational Algebraic Geometry and Commutative Algebra, Springer-Verlag, New York, 1992.
[12] J. Stoer, C. Witzgall, Convexity and Optimization in Finite Dimensions I, Springer, Berlin, 1970.
[13] P. Veelaert, K. Teelen, Fast polynomial segmentation of digitized curves, in: Lecture Notes in Computer Science: Proc. Discrete Geometry for Computer Imagery, vol. 4245, Springer, Szeged, 2006, pp. 482–493.